# On the Nature of Bayesian Convergence[1]

James Hawthorne

University of Oklahoma

## 1. Introduction

Bayesians assess the inductive support for theoretical hypotheses on the basis of two sorts of factors, one fairly objective, the other highly subjective. The objective factor consists of the *likelihoods* or *direct inference probabilities* that theoretical hypotheses specify for evidential events. This is the means by which evidence affects inductive support. The subjective factor consists of the *prior probabilities* assigned to the various competing hypotheses. For a Bayesian agent the *prior* probability of a hypothesis represents how plausible the agent considers the hypothesis to be *before* the impact of evidence is considered, and Bayesian agents may radically differ in their initial plausibility assessments. Bayes' formula combines likelihoods with an agent's prior probabilities to produce the agent's *posterior probability* for each hypothesis. *Posterior* probabilities represent how plausible an agent considers the hypothesis to be *after* the evidence is taken into account. Thus, to the extent that the values of subjective prior probabilities continue to affect the values of posterior probabilities as evidence accumulates, Bayesian agents will continue to differ regarding how plausible they consider theoretical hypotheses to be. If Bayesian induction is to yield either an objective assessment or intersubjective agreement among agents regarding the inductive support for hypotheses, then the evidence must somehow produce a *convergence to agreement* among the posterior probabilities of different Bayesian agents in spite of their initial disagreements about the plausibilities of the hypotheses. Hence, advocates of Bayesian induction have investigated the circumstances under which evidence might "wash out" the effects of subjective priors and bring agents into agreement on posterior probabilities. Bayesian convergence theorems establish conditions under which accumulating evidence can induce agents to converge to agreement.

In this paper I will describe some important features of Bayesian convergence, including some features that have not been widely recognized by Bayesian logicians. I will discuss three sorts of Bayesian convergence results. The first shows how the objectivity of simple inductions, which assess the likelihoods of individual events, depends on the objectivity of posterior probabilities of general theoretical hypotheses. This sort of convergence result shows the pivotal role of theoretical hypotheses in *systematizing* simple inductive inferences in a manner that bolsters their objectivity.

PSA 1994, Volume 1, pp. 241-249
Copyright © 1994 by the Philosophy of Science Association

The second convergence result reveals that (except in very special circumstances) evidence can induce Bayesian probability functions to *converge to agreement* on the posterior probabilities for theoretical hypotheses *only if* the convergence is to 0 (refutation) or 1 (confirmation). The third result establishes general conditions under which the evidence *will very probably* compel posterior probabilities of theoretical hypotheses to converge to 0 or 1.

## 2. Simple Induction

*Simple induction* is a form of inference in which an agent infers the probability that some particular outcome e will result from some initial state of affairs c on the basis of some sequence of evidential events. Simple inductions are inferences about specific "occurrent events" rather than general hypotheses or theories. Let '$(c_1 \cdot c_2 \cdot ... \cdot c_n)$' represent a conjunction of descriptions of the initial states, initial conditions, or experimental arrangements for a series of past observations, and let their respective outcomes form the conjunction '$(e_1 \cdot e_2 \cdot ... \cdot e_n)$'. I will use '$c^n$' to abbreviate the conjunction '$(c_1 \cdot c_2 \cdot ... \cdot c_n)$' of descriptions of n initial states, and '$e^n$' to abbreviate the conjunction of the n descriptions of respective outcomes. Agents may also employ some relatively uncontroversial background knowledge b as a premise in simple inductive inferences. Background knowledge (including auxiliary hypotheses) represented by 'b' will typically include relatively uncontroversial claims about how pieces of familiar instrumentation work and about methods and conditions under which various kinds of observable phenomena may be reliably detected or measured.

The evidence for a simple induction may be a sequence of events that are very similar to the event whose probability is to be inferred—as when we infer the probability that the next toss of a particular bent coin will come up heads from a sequence of outcomes of previous tosses of the same coin. In such cases simple induction is typically called *induction by enumeration*. But, under the term *simple induction* I mean to include a much broader class of inferences than those that merely employ enumerations of similar cases. In general the evidence for simple inductions may be a very diverse collection of previous events. The evidence might include descriptions of how other objects of various shapes have tumbled, and how they have bounced on different surfaces. Indeed, the simple inductive evidence for an event could include all of the evidence that one might normally employ in the confirmation of the kinds of theoretical hypotheses that would be relevant to the individual event in question.

Let $\alpha$ and $\beta$ be two Bayesian agents whose respective probability functions are $P_\alpha$ and $P_\beta$. For any sentences A and B, $P_\alpha[A \mid B]$ represents $\alpha$'s conditional probability for A given B, her assessment of how probable (how plausible, how likely to be true) A would be if B were true (and similarly for $P_\beta[A \mid B]$). For r a real number between 0 and 1, '$P_\alpha[e \mid c \cdot c^n \cdot e^n \cdot b] = r$' represents the *simple inductive probability* of e on $(c \cdot c^n \cdot e^n \cdot b)$ for agent $\alpha$. I.e. if the premises in the conjunction '$(c \cdot c^n \cdot e^n \cdot b)$' represent $\alpha$'s total relevant evidence regarding e, then $\alpha$'s rational degree of confidence in e, the plausibility of e for $\alpha$, should be r.

Even when agents possess the same total evidence, they may strongly disagree on the simple inductive probability of an outcome e of condition c—i.e. the probability values r and s for the simple inductive inferences $P_\alpha[e \mid c \cdot c^n \cdot e^n \cdot b] = r$ and $P_\beta[e \mid c \cdot c^n \cdot e^n \cdot b] = s$ may differ greatly. How might agents be induced to come to agreement about the probability of e as the evidence accumulates? If the evidence could bring the agents to concur on the truth of some general theoretical hypothesis that states deterministic or stochastic laws governing situations like c, then they would come to agree on how likely e is. Each agent would invoke the theoretical hypothesis, together with c and b, and arrive at the same conclusion about the likelihood that e is true. Indeed, one of the primary reasons that agents seek to confirm a theoretical hypothesis is so that it may be used as a kind of objective inference ticket that *systematizes simple inductions*. The role of theoretical hypotheses in the systematization of simple inductions is central to the scientific enterprize. We will see precisely how inductive systematization works in a Bayesian context after first exploring the relationship between theoretical hypotheses and descriptions of individual events from a Bayesian perspective.

## 3. Theoretical Hypotheses and Likelihoods

Let $H = \{h_1, h_2, ...\}$ be a class of competing theoretical hypotheses that bear on whether event e will result from condition c. H may contain an infinite number of alternatives, and two Bayesian agents $\alpha$ and $\beta$ may disagree widely on how plausible they are. But suppose that $\alpha$ and $\beta$ are like-minded enough that they consider only the theoretical hypotheses in H to have some non-zero degree of initial plausibility. Thus, for each agent the sum of prior probabilities of hypotheses in H is 1 (i.e. $\Sigma_j P_\alpha[h_j|b] = 1 = \Sigma_j P_\beta[h_j|b]$). And since the hypotheses in H are alternatives, any distinct pair of them, $h_i$ and $h_j$, should be incompatible (i.e. $P_\alpha[h_i \cdot h_j|b] = 0$, and similarly for $\beta$).

The hypotheses in H may be deterministic or statistical, and they may be extremely broad theories or their scope may be quite narrow. But whatever their scope, Bayesians typically suppose that the principal epistemic role of theoretical hypotheses is to underwrite *relatively objective* probabilities for individual events. Taken together with initial conditions c and background knowledge b, each theoretical hypothesis $h_i$ should provide a fairly unambiguous indication of the probability that e will occur (if $h_i$ is true). This is one of the main reasons that we construct theoretical hypotheses in the first place. Bayesians usually call these probabilities *likelihoods* or *direct inference probabilities*. Likelihoods take the form $P_\alpha[e|c \cdot h_i \cdot b] = r$, and Bayesian agents are supposed to generally agree on their values.

*Logicist Bayesians* usually call likelihoods *direct inference* probabilities. Logicists think of likelihoods as objective, logical relationships, as an extension of the logical entailment relation to a form of probabilistic entailment that can logically link stochastic object-language sentences (e.g. about propensities) to descriptions of individual chance events. By contrast, *Personalist Bayesians* do not think of likelihoods as *logically determinate*, but they also usually consider likelihoods to have a "more objective" status than other probabilities. Personalists often maintain that relative to theoretical hypotheses there will usually be a high degree of intersubjective agreement among Bayesian agents on the likelihoods of events. Indeed, the main rational for using Bayes' Theorem to *calculate posterior probabilities* of theoretical hypotheses is that the likelihoods are generally considered by Bayesians to be more objective, or more subject to intersubjective agreement than the posterior probabilities they are used to calculate. Henceforth I will suppose that agents agree on likelihoods for events relative to hypotheses in H. As a result we may drop the subscripts '$\alpha$' and '$\beta$' from likelihoods, and for each hypothesis $h_i$ we write $P_\alpha[e|c \cdot h_i \cdot b] = P_\beta[e|c \cdot h_i \cdot b] = P[e|c \cdot h_i \cdot b]$.

There is another important facet to the kind of objectivity that theoretical hypotheses should afford likelihoods of events. Consider the probability of event e relative to both $(c \cdot h_i \cdot b)$ *and* to the previous evidence $(c^n \cdot e^n)$—i.e. consider $P_\alpha[e|(c \cdot h_i \cdot b) \cdot (c^n \cdot e^n)]$. When $(c \cdot h_i \cdot b)$ furnishes a direct inference probability $P[e|c \cdot h_i \cdot b] = r$, the old evidence $(c^n \cdot e^n)$ should become irrelevant to the probability of e (relative to $(c \cdot h_i \cdot b)$). The idea is that the old evidence plays it's role through the inductive support it provides for $h_i$. But given that $h_i$ (together with $(c \cdot b)$) *is* true, the old evidence should be screened off from influence on the probability of e. Thus, we may reasonably suppose that the following *independence condition* holds for each Bayesian agent $\alpha$: $P_\alpha[e|(c\cdot h_i\cdot b)\cdot(c^n\cdot e^n)] = P[e|c\cdot h_i\cdot b]$. If a hypothesis failed to *screen off* predicted events from previous evidence in this way, then each time an agent appealed to the hypothesis to predict an event she would have to employ a vast collection of previous observational and experimental data as initial conditions. This would largely undermine one of the chief reasons for formulating theoretical hypotheses in the first place. Imagine what it would be like if to compute a future location for a planet we not only had to appeal to a gravitational theory and a few observations of the planets' past locations, but also had to employ in the computation all of the data from the experiments and observations that went to confirm the theory of gravitation.

Finally, another plausible *independence condition* will be used in the next section. The initial condition statement c that combines with a hypothesis to determine the likelihood of an outcome should not by itself function as evidence for or against the hypothesis. It should only become relevant to the support of a hypothesis when it is conjoined with its associated outcome e. That is, although adding (c·e) to any previous evidence might count as additional evidence for $h_i$, when the outcome e is still in question, c should not by itself be relevant to the posterior probability of $h_i$. Thus, for each agent $\alpha$ we should have $P_\alpha[h_i|c\cdot c^n\cdot e^n\cdot b] = P_\alpha[h_i|c^n\cdot e^n\cdot b]$. (Equivalently, relative to $(c^n\cdot e^n\cdot b)$, the initial condition c should be no more likely to have occurred if $h_i$ is true than if it is false, i.e. $P\alpha[c|h_i\cdot(c^n\cdot e^n\cdot b)] = P\alpha[c|\neg h_i\cdot(c^n\cdot e^n\cdot b)]$.)

We are now ready to see how agents may be induced to come to agreement about the probability of an event e as the evidence accumulates. The main idea is simple. If the accumulating evidence comes to strongly support the same hypothesis in H for all agents, then that hypothesis may be used as a premise that generates a likelihood for e on which all agents agree. Thus, inductive confirmation of hypotheses can systematize simple inductive inferences. In the next section I will spell out the details of this strategy from a Bayesian perspective.

4. Bayesian Convergence for Simple Inductions

Under the conditions described in the previous section the convergence to agreement of simple inductions for Bayesian agents reduces to their convergence to agreement on the posterior probabilities of the theoretical hypotheses in H. To see this notice that for any Bayesian agent $\alpha$:

$$\text{(I)} \quad P_\alpha[e|c\cdot c^n\cdot e^n\cdot b] = \Sigma_j P_\alpha[e|(c\cdot h_j\cdot b)\cdot(c^n\cdot e^n)] \times P_\alpha[h_j|c\cdot c^n\cdot e^n\cdot b] = \Sigma_j P[e|c\cdot h_j\cdot b] \times P_\alpha[h_j|c^n\cdot e^n\cdot b].$$

The first line of this equality is a theorem of probability theory (since H is a set of mutually incompatible hypotheses to which Bayesian agents assign probabilities that sum to 1). The second line follows directly from the first line and the two independence conditions described in the previous section.

Equation (I) exhibits the connection between simple inductions, likelihoods, and Bayesian theory confirmation. It suggests two different sorts of convergence for simple inductions. The first is the kind that comes from highly confirming a hypothesis in H. If the posterior probability that $\alpha$ assigns hypothesis $h_i$ approaches 1 as the evidence increases (i.e. if $P_\alpha[h_i|c^n\cdot e^n\cdot b] \longrightarrow 1$, as n increases), then the simple induction probabilities for $\alpha$ will approach the likelihoods that $h_i$ specifies (i.e. $P_\alpha[e|c^n\cdot e^n\cdot b] \longrightarrow P[e|c\cdot h_i\cdot b]$). Call the convergence to 1 of agent $\alpha$'s posterior probabilities for a hypothesis (and the convergence to 0 of its alternatives) *Type 1 Hypothesis Convergence* for $\alpha$.

Call the convergence of $\alpha$'s simple inductive probabilities to the likelihoods specified by a hypothesis *Type 1 Simple Inductive Convergence* for $\alpha$. Then the first convergence result that flows from equation (I) says that, for each Bayesian agent:

*Type 1 Hypothesis Convergence* implies *Type 1 Simple Inductive Convergence.*

Equation (I) also suggests a second sort of Bayesian convergence, a variety of convergence to agreement *among* agents. Call the convergence to agreement of agents $\alpha$ and $\beta$ on their posterior probabilities for all hypotheses in H *Type 2 Hypothesis Convergence* for $\alpha$ and $\beta$ (i.e. for each $h_j$, $| P_\alpha[h_j|c^n \cdot e^n \cdot b] - P_\beta[h_j|c^n \cdot e^n \cdot b] | \longrightarrow 0$). Call the convergence to agreement by $\alpha$ and $\beta$ on the probabilities for simple inductions *Type 2 Simple Inductive Convergence* for $\alpha$ and $\beta$ (i.e. $| P_\alpha[e|c^n \cdot e^n \cdot b] - P_\beta[e|c^n \cdot e^n \cdot b] | \longrightarrow 0$). Then the second sort of convergence result that flows from equation (I) is that, for each pair of agents:

*Type 2 Hypothesis Convergence* implies *Type 2 Simple Inductive Convergence.*

Each of the two types of *simple inductive convergence* lends a kind of objectivity to simple Bayesian inductions. Each shows how the Bayesian evaluation of hypotheses leads to a systematization of simple inductions by reducing the convergence problem for simple inductions to a convergence problem for theoretical hypotheses. Indeed, inductive systematization turns out to be even more regimented than one might have expected. For, somewhat surprisingly:

*Type 2 Hypothesis Convergence* implies *Type 1 Hypothesis Convergence.*

That is, if evidence can bring a pair of Bayesian agents who possess even *moderately diverse* prior probabilities for theoretical hypotheses (in a sense to be made precise in the next section) into agreement about posterior probabilities, then they must come to agree that the posterior probability of one particular hypothesis, $h_i$, approaches 1 (and posteriors of its competitors approach 0).

In the next section I will explain the connection between the two types of hypothesis convergence in more detail. Its implications for the nature of simple induction are striking. If evidence induces agreement among Bayesian agents on the probability of an event (*Type 2 Simple inductive Convergence*) by causing the agents to converge to agreement on the posterior probabilities of hypotheses (*Type 2 Hypothesis Convergence*), then it may do so *only* by raising the posterior probability of one hypothesis towards 1 for each Bayesian agent (*Type 1 Hypothesis Convergence*), thus forcing the simple inductions of all agents towards agreement with the (direct inference) likelihoods specified by that hypothesis (*Type 1 Simple Induction Convergence*).

## 5. Likelihood Ratios and Bayesian Convergence for Theories

Bayesian induction regarding theoretical hypotheses essentially depends on *likelihood ratios*. Likelihood ratios are ratios of direct inference probabilities for competing hypotheses and will be abbreviated by '$LR[e^n|j/i]$', where by definition:

$$LR[e^n|j/i] = P[e^n|c^n \cdot h_j \cdot b] / P[e^n|c^n \cdot h_i \cdot b].$$

Likelihood ratios measure how many times more (or less) likely the evidence would be according to one hypothesis as compared to another.

The central role of likelihood ratios in Bayesian induction becomes apparent when we consider the ratio of posterior probabilities for a pair of hypotheses. The ratio of their posterior probabilities equals the product of their likelihood ratio with the ratio of their prior probabilities:

(II) $P_\alpha[h_j|e^n \cdot c^n \cdot b] / P_\alpha[h_i|e^n \cdot c^n \cdot b] = LR[e^n|j/i] \times (P_\alpha[h_j|b] / P_\alpha[h_i|b])$.

This equality is a theorem of probability theory (provided that the probability of the initial conditions is the same relative to each hypotheses— i.e.$P_\alpha[c^n|h_i \cdot b] = P_\beta[c^n|h_j \cdot b]$). For simplicity I will assume that this proviso holds; a much weaker assumption would suffice, see (Hawthorne, forthcoming), but would complicate the exposition unnecessarily.

The relationship between ratios of posterior probabilities and likelihood ratios expressed by equation (II) is really all there is to Bayesian induction for theoretical hypotheses. The *absolute* probability of a hypothesis comes directly from the sum of these ratios. To see this, first consider the *odds*, $\Omega_\alpha$, against a hypothesis $h_i$ relative to evidence, defined as follows:

(III) $\Omega_\alpha[\neg h_i|e^n \cdot c^n \cdot b] = P_\alpha[\neg h_i|e^n \cdot c^n \cdot b] / P_\alpha[h_i|e^n \cdot c^n \cdot b]$
$= \sum_{j \neq i} P_\alpha[h_j|e^n \cdot c^n \cdot b] / P_\alpha[h_i|e^n \cdot c^n \cdot b]$.

The odds against a hypotheses is the sum of the relative plausibilities for its competitors, the sum of instances of equation (II). Equations (II) and (III), then, imply the following relationship between the odds against a hypothesis and likelihood ratios:

(IV) $\Omega_\alpha[\neg h_i|e^n \cdot c^n \cdot b] = \sum_{j \neq i} LR[e^n|j/i] \times (P_\alpha[h_j|b] / P_\alpha[h_i|b])$.

The probability of a hypothesis on evidence is related to the odds against the hypothesis by the following formula:

(V) $P_\alpha[h_i|e^n \cdot c^n \cdot b] = 1 / (1 + \Omega_\alpha[\neg h_i|e^n \cdot c^n \cdot b])$.

Taken together, equations (IV) and (V) express a form of Bayes's theorem in terms of the odds against a hypothesis; this formulation makes the role of the likelihood ratios more perspicuous than the more usual form of the theorem, which is:

(VI) $P_\alpha[h_i|e^n \cdot c^n \cdot b] = P[e^n|c^n \cdot h_i \cdot b] \times P_\alpha[h_i|b] / \sum_j P[e^n|c^n \cdot h_j \cdot b] \times P_\alpha[h_j|b]$.

If $h_i$ comes to make the evidence negligibly likely in comparison to some alternative, $h_j$—i.e. if $LR[e^n|i/j]$ converges to 0—then the inverse likelihood ratio, $LR[e^n|j/i]$, blows up to infinity, and the odds against $h_i$ blow up with it, by equation (IV). Hence, by equation (V), the probability of $h_i$ goes to 0. On the other hand, if *every* alternative to $h_i$ makes the evidence negligibly likely in comparison to $h_i$—i.e. if for every alternative $h_j$, $LR[e^n|j/i]$ converges to 0—then by equation (IV) the odds against $h_i$ converge to 0. When this happens, equation (V) says that the probability of $h_i$ converges to 1.

Now, suppose the accumulating evidence does not drive the probability of $h_i$ to either 0 or 1. Then for some alternative hypothesis $h_j$, the likelihood ratios $LR[e^n|j/i]$ will neither blow up nor converge to 0. If these likelihood ratios do not go to extremes, then equation (II) implies that the ratio of posterior probabilities of $h_j$ to $h_i$ will remain under the influence of their prior probabilities. Thus, if the prior probabilities of agents $\alpha$ and $\beta$ diverge radically for $h_i$ and $h_j$, then so must their posterior probabilities. Indeed, if the likelihood of evidence relative to $h_i$ agrees with the likelihood relative to $h_j$, then $LR[e^n|j/i] = 1$, and the evidence yields *no change* in the ratio of their posterior probabilities from the ratio of their priors. So, when the evidence fails to take the likelihood ratios to extremes, the initial plausibility assessments will continue to significantly affect the posterior probabilities of hypotheses.

In light of the central role played by likelihood ratios in Bayesian induction the following theorem should not be too surprising. See (Hawthorne, forthcoming) for details and a proof.

Theorem. *Non-Zero Convergence is Convergence to One.*
Let $h_i$ be some hypothesis in H, and suppose that the following conditions are satisfied:
i) for agent $\alpha$ there is a number r such that, for all n, $P_\alpha[h_i|e^n \cdot c^n \cdot b] \geq r > 0$;
ii) there is another agent $\beta$ who's probability function $P_\beta$ *modestly differs* with $P_\alpha$ on the prior plausibilities for hypotheses in the sense that there is some fraction q between 0 and 1 such that, for every $h_j$ in H other than $h_i$, $P_\beta[h_j|b] \leq q \times P_\alpha[h_j|b]$;
iii) $\lim_n | P_\alpha[h_i|e^n \cdot c^n \cdot b] - P_\beta[h_i|e^n \cdot c^n \cdot b] | = 0$.
*Then,* for every $h_j$ in H other than $h_i$, $\lim_n LR[e^n|j/i] = 0$; and $\lim_n P_\alpha[h_i|e^n \cdot c^n \cdot b] = 1$.

A *modest difference* between two agents $\alpha$ and $\beta$ (as expressed in condition (ii)) simply means that for all alternatives to a hypothesis $h_i$ in H, agent $\beta$'s prior probabilities for the alternatives are at least slightly below (e.g. less than 99.99% of) the respective prior probabilities that $\alpha$ assigns them. The only hypothesis to which $\beta$ assigns a higher prior probability than $\alpha$ is $h_i$. The theorem implies that if a community of Bayesian agents is diverse enough to contain even one pair of agents who *modestly differ* on a hypothesis $h_i$ (which doesn't acquire an arbitrarily low posterior probability), then these two agents can *converge to agreement* about the posterior probabilities of hypotheses *only if* the whole community comes to agree that the evidence increasing confirms $h_i$ and increasing refutes its competitors (since, for all alternatives $h_j$, $\lim_n LR[e^n|j/i] = 0$ *for every agent*). (Alternatively, if $h_i$ acquires an arbitrarily low posterior probability for some agent $\alpha$, then for some $h_j$, $LR[e^n|j/i]$ must get arbitrarily large, by equations (V) and (IV); so $h_i$ acquires an arbitrarily low posterior probability for *all* agents.)

It takes just one *modestly differing* pair of agents in the community for the theorem to apply. There are, of course, special classes of probability functions, representing highly restricted communities of Bayesian agents, for which convergence short of 0 and 1 may occur. Suppose, for example, that all agents in a community agree on the prior probabilities for most hypotheses, but disagree on priors of a few. If all of the hypotheses on which they initially disagree become increasingly refuted by the evidence, then everyone in the community will converge on common values other than 0 or 1 for posterior probabilities of the unrefuted hypotheses. However, if the community contains even one agent who's probability function *modestly differs* from another's (on a hypothesis who's posterior remains above some $r > 0$), then convergence to agreement implies that every hypothesis converges to 0 or 1 *for all agents in the community.*

Finally, notice that evidence can only distinguish between hypotheses in H that disagree on the likelihoods of at least some possible events. The influence of the prior probabilities of empirically equivalent hypotheses cannot be washed out unless each is refuted relative to some other empirically distinct hypothesis. If empirically equivalent alternatives to the true hypothesis are among the alternatives in H, then at best the evidence can refute their empirically distinct competitors, and bring the disjunction of the true hypothesis with its empirical equivalents in H to converge to 1. The relative sizes of posterior probabilities of the true hypothesis and its empirical equivalents remains equal to the relative sizes of their priors (by equation (II)).

## 6. The Likelihood of Obtaining Refuting Evidence

Thus far I have argued that Bayesian convergence ultimately reduces to the convergence of posterior probabilities of theoretical hypotheses to either 0 or 1. Equations (IV) and (V) together show that a hypothesis $h_i$ can become highly refuted only if at least one alternative $h_j$ makes the evidence much more likely than does $h_i$, so that LR[$e_n$|j/i] blows up. And $h_j$ may in turn become refuted relative to some other hypothesis. If, however, $h_i$ is to become highly confirmed, it can only do so only by driving LR[$e_n$|j/i] to 0 for all alternative hypotheses $h_j$ in H. Thus, the crucial question becomes: Is there any reason to think that accumulating evidence will cause empirically distinct alternatives of the true hypothesis to become *increasingly refuted* relative to the true hypothesis via likelihood ratios? I will briefly describe a third kind of Bayesian convergence theorem that establishes that if two hypotheses are empirically distinct, then a sufficiently long sequence of evidence can almost certainly do the job.

L.J. Savage's Bayesian convergence theorem (1972, 46-50) is just such a result. It says that if the accumulating evidence consists of a sequence of independent, identically distributed events (i.e. if the evidence is drawn from repetitions of the same kind of observation or experiment, like repeated tosses of the same coin), then false alternative hypotheses will *almost certainly* become *highly refuted*, and the true hypothesis (or its disjunction with empirically equivalent competitors) will become *highly confirmed*.

Hesse (1975) and Earman (1992) argue convincingly that Savage's theorem presupposes conditions on the evidence that are unrealistic for most real cases of scientific theory testing. In particular, Savage's assumption that the evidence is a sequence of independent, identically distributed events is generally not satisfied. Hesse and Earman also point out that Savage's theorem puts no bounds on the rate at which convergence takes place; for all we know a billion observations would not be enough to bring about any noticeable degree of convergence.

There is a generalized version of Savage's theorem, (in Hawthorne, forthcoming) that avoids the main objections raised by Hesse and Earman. In this version of the theorem the evidence need not consist of identically distributed events, nor is it required to consist of independent events (although the independence of evidential events *relative to a theory* is a perfectly reasonable assumption in scientific contexts). This version also provides bounds on the rate of Bayesian convergence that explicitly depend on a quantitative information-theoretic measure of the *quality of the evidence*. I do not have space to go deeply into the details of the theorem, but I will briefly describe its main features.

Suppose that $h_i$ is a true hypothesis in H (although, of course, the agents are unaware of its truth) and let $h_j$ be one of its empirically distinct competitors. Also suppose that the background claims b are true and that a sequence of initial states or observational conditions $c^n$ holds. Let $E^n$ be the set of all possible outcome sequences that might result from $c^n$. That is, each member of $E^n$ is a conjunction $e^n$ that describes a possible sequence of outcomes that may result from the conditions described by $c^n$.

Hypothesis $h_i$ will assign a higher likelihood to some of the outcome sequences described in $E^n$ than does the competitor $h_j$, and it will assign a lower likelihood to others. This is just what it means for $h_i$ and $h_j$ to be empirically distinct. Consider the following subset of possible outcome sequences in $E^n$: $S^n(m) = \{e^n \mid e^n \in E^n$ and $LR[e^n|j/i] < 1/2^m\}$. Each outcome sequence in the set $S^n(m)$ is one for which the likelihood according to $h_i$ is a factor of more than $2^m$ (e.g. $2^{100}$) times larger than the likelihood $h_j$ specifies for it.

Let '$V[S^n(m)]$' denote the *disjunction* of all of the possible outcome sequences in $S^n(m)$. '$V[S^n(m)]$' asserts that, for the first n observations $c^n$, one of the outcome sequences will occur that makes the likelihood ratio $LR[e^n|j/i]$ less than $1/2^m$. If '$V[S^n(m)]$' *is true* for some very large value of m (e.g. m = 100), then, an $e^n$ *does occur* that makes $LR[e^n|j/i]$ extremely small (e.g. less that $1/2^{100}$). Generally this will make $h_j$ highly unlikely on the evidence (relative to $h_i$).

How likely is it that '$V[S^n(m)]$' is true? The true hypotheses $h_i$ answers this question with the following *direct inference probability*:

$$P[\ V[S^n(m)] \mid c^n \cdot h_i \cdot b] \ \geq\ 1 - (1/n) \times \underline{VQI}^n[i/j|i] \,/\, (\underline{EQI}^n[i/j|i] - (m/n))^2.$$

This probability will converge to 1 as n increases, provided only that $h_j$ differs from $h_i$ (at least slightly) about the likelihoods of some possible outcome sequences. (The terms $\underline{EQI}^n[i/j|i]$ and $\underline{VQI}^n[i/j|i]$ are information-theoretic measures of the extent of disagreement between $h_i$ and $h_j$ about the likelihoods of the various possible outcomes. They are, respectively, a measure of the *mean* and *variance* of the *expected values* of $\log(LR[e^n|i/j])$ in the set $E^n$. If $h_j$ differs even slightly from $h_i$ regarding the likelihoods of some possible outcome sequences in $E^n$, for increasing values of n, then the ratios $(\underline{VQI}^n[i/j|i] \,/\, \underline{EQI}^n[i/j|i]^2)$ will be bounded above. It follows immediately that for any chosen value of m, $P[\ V[S^n(m)] \mid c^n \cdot h_i \cdot b]$ goes to 1 as the amount of evidence, n, increases. See (Hawthorne, forthcoming) for details.

Thus, if $h_i$ is true, then for each empirically distinct alternative hypothesis $h_j$, the likelihood ratio $LR[e^n|j/i]$ will almost surely go to 0. We saw earlier that this kind of convergence brings with it a more general convergence, the convergence of simple inductive inferences to the values of direct inference probabilities. Therefore, if Bayesian agents discover a true hypothesis and test empirically distinct competitors against it in a contest of likelihood ratios, they will almost surely come to agree on the inductive support for theories and on the simple inductive probabilities of individual events.

## Notes

[1] I wish to thank Chris Swoyer for numerous valuable comments and suggestions.

## References

Earman, J. (1992), *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge: MIT Press.

Hawthorne, J. (forthcoming), "Bayesian Induction Is Eliminative Induction", *Philosophical Topics*, v. 21, no. 1.

Hesse, M. (1975), "Bayesian Methods and the Initial Probability of Theories", in G. Maxwell and R. Anderson, (eds.), *Induction, Probability, and Confirmation*, Minnesota Studies in the Philosophy of Science, vol. 6. Minneapolis: U. of Minnesota Press, pp. 50-105.

Savage, L. (1972), *The Foundations of Statistics*. New York: Dover.